



Managed by  Duke Clinical & Translational Science Institute

The MURDOCK Study Community Registry and Biorepository is a 12,526-participant community-based longitudinal cohort recruited from a 20-Zip Code region in the Southeastern United States (U.S.) that is centered in the city of Kannapolis, NC and encompasses Cabarrus County, NC.

Creation of the cohort was funded by a gift to Duke University from the David H. Murdock Institute for Business and Culture, with operational support from Duke's Clinical and Translational Science Award (CTSA) grant (UL1TR002553) and the Duke Clinical and Translational Science Institute (CTSI).

Consenting participants complete a baseline health questionnaire at enrollment, as well as a brief physical exam and collection of blood and urine. Consent includes permission to access to information from medical records, storage of collected samples in the biorepository, access to collected data and biospecimens for future approved research studies and contact regarding new research study opportunities.

Data have been organized into "storefronts" that summarize characteristics of a population of research interest as well as available data and samples for that population. The following sections summarize the sources of data in the MURDOCK Study database, as well as important descriptions and definitions to help understand the data presented in the "storefronts".

**1 Participant self-reported data at baseline.** The baseline questionnaire collects contact information, current residential street address, and primary physician; alternate contact information; date and place of birth; demographics; current or past diagnosis of 34 medical conditions; menopausal status in women; medications, vitamins and supplements; dietary and physical activity assessment; hours of sleep per night; tobacco and alcohol use; second-hand smoke exposure; and selected PROMIS® participant-reported outcomes domains. Socioeconomic data collected at baseline included marital status, highest level of education of participant and participant's mother and father, employment status, mother's and father's occupations, housing (type, how paid for, number of adults and children in the household) and total household income. In addition, a brief physical exam (vital signs, height, weight, and waist circumference) was conducted at enrollment.

**Medical conditions:** "Do you have, or have you ever had, any of the following [medical conditions]?" (yes, no, don't know). Counts are unique participants reporting yes to specific condition. **Medications:** "Please list any pharmaceutical and/or natural medications (including vitamins) that you are currently taking." Data are captured in free-text format as written by the participant and coded using RxNorm. Summary metrics are based on everything reported. Top 5 reported medications are limited to reported prescriptions.

**2 Biorepository samples.** Blood was collected at baseline and processed into the following specific samples: whole blood in EDTA for DNA extraction, whole blood in PAXgene for RNA extraction, plasma, serum and buffy coat in cryovials. Urine was collected and aliquoted in cryovials. Serial sample collection was not done systematically for MURDOCK enrollees; however, some nested sub cohorts and other studies enrolling MURDOCK registry participants include sample collection at follow up time points. All samples are stored at -80°C in a central biorepository currently managed by Fisher BioServices, a division of Thermo Fisher Scientific, under a contractual agreement with Duke University.

**Samples in inventory:** Data are summarized by sample type as well as specific container and size. Participant counts are unique individuals with one or more aliquots. Aliquot counts are all unique samples for a given type and container, size. Freezers is a calculation of approximate storage requirements based on sample type/size, box size, and number of boxes that can be stored per freezer.

**3 Participant self-reported changes in health via annual follow up.** Participants are asked to complete a follow-up form once a year around the time of their original enrollment date. Participants may update contact information, primary care physician/practice and alternate contact. PROMIS domains are repeated at each annual time point in order to capture changes in participant-reported outcomes over time. The form collects new incidence/diagnosis of the same 34 medical conditions surveyed at baseline. Hospitalizations during the past year are collected along with reason, as well as specific medical procedures. Participants may update their medication list to reflect current medications, vitamins and supplements being taken at the time of follow up form completion.

**Vital status:** Death reported by family member or alternate contact is confirmed by obituary as the primary source. Cause of death is not captured. **Follow-up metrics:** Follow-up is defined as complete if participant fills out the survey online or by mail or phone. Completeness is measured as surveys completed relative to years eligible to complete follow-up. **Medical conditions:** "Please indicate if you have received a new diagnosis of any of the following medical conditions in the past year (yes, no, don't know)". Counts and percentages are unique participants reporting yes to specific condition in follow-up for participants that did NOT report yes at baseline. **Procedures:** "Please indicate if you have any of the following medical procedures in the past year". Counts are unique participants reporting the specified procedure one or more times during follow up. **Hospitalizations:** Participants are asked to report if they have been hospitalized within the last year, for each hospitalization they are asked to list reason(s) for hospitalization, admission date and hospital name. Reasons for hospitalization are captured as free-text responses as written by participants. Responses are coded, when possible, in order to list the most frequently reported reasons for hospitalization. **Medications:** (see note above for medications reported at baseline). The denominator for data based on last follow-up are participants with at least one follow-up survey complete.

**4 Electronic health record (EHR) data from regional healthcare providers.** Duke has partnered with regional healthcare providers to integrate data from EHR systems for consented MURDOCK Study participants. Participants are identified in EHR systems with robust matching algorithms using common identifiers from the MURDOCK and EHR databases. Data are transferred under a data use agreement (DUA) with the specific provider organization which specifies the scope of data and frequency of transfers. Data availability vary by participant and depend on whether or not a participant has had one or more encounters with the healthcare provider system during the time period included in the dataset.

**Available EHR datasets:** Data are summarized by healthcare provider organizations. Counts are unique participants with one or more ICD codes in the EHR dataset. **Available EHR domains:** Data are summarized by domain in the EHR dataset. Counts are unique participants with one or more records (rows of data) for the specified domain. **Insights from available EHR data:** Specific EHR data related to the population of research interest is presented with granularity when possible.

**5 Additional data collection from studies with MURDOCK participants.** MURDOCK Study participants may be recruited to enroll in additional research study opportunities by Duke researchers or other collaborators. Data sharing is a condition of collaboration with the MURDOCK Study; therefore, data collected from MURDOCK Study participants and/or generated from biospecimens as part of additional research studies is returned for integration with all other MURDOCK registry data.

"Storefronts" for nested sub-cohorts summarize surveys, assessments and/or other data collected specifically as part of enrollment and participation in the study. **Samples in inventory:** Samples are summarized if collected (see note above for samples collected at baseline). **Participation in other studies:** Counts are participants from the population of research interest enrolled in the specified study listed. *Brief descriptions of relevant studies are listed along with a summary of study procedures and/or data collected.*

**MURDOCK Study participants with rheumatoid arthritis, N=1,712**
**Participant self-reported characteristics at MURDOCK Study enrollment (baseline, February 2009 – February 2018)**

Demographics at baseline		Education at baseline				
<b>Age</b>	<b>Baseline</b>	Less than high school graduate	265 (15%)			
Median (25 <sup>th</sup> , 75 <sup>th</sup> )	60 (50, 68)	High school graduate, equivalent	467 (27%)			
Min, Max	<18, 90+	Some college or associates degree	636 (37%)			
<b>Sex</b>		Bachelor's degree	219 (13%)			
Female	1,148 (67%)	Master's or higher professional degree	124 (7%)			
Male	564 (33%)	<b>Income at baseline</b>				
<b>Race</b>		Under \$10,000	196 (11%)			
American Indian & Alaska Native	7 (<1%)	\$10,000-29,999	435 (25%)			
Asian	6 (<1%)	\$30,000-49,999	308 (18%)			
Black or African American	335 (20%)	\$50,000-69,999	210 (12%)			
Native Hawaiian & Other Pacific Islander	1 (<1%)	\$70,000-89,999	135 (8%)			
White/Caucasian	1,180 (69%)	\$90,000 or more	178 (10%)			
Other	128 (7%)	Don't know, no response	250 (15%)			
Multiple	40 (2%)	<b>Body mass index (BMI) at baseline</b>				
Don't know/Not sure/Not answered	15 (1%)	<18.5 (underweight)	18 (1%)			
<b>Ethnicity</b>		18.5 - 24.9 (normal weight)	308 (18%)			
Hispanic or Latino	173 (10%)	25 - 29.9 (overweight)	565 (33%)			
Non-Hispanic or Latino	1,500 (88%)	30+ (obese)	809 (48%)			
Don't know/Not sure/Not answered	39 (2%)	<b>Exercise at baseline</b>				
<b>Smoking history at baseline</b>		Little to no physical activity	846 (49%)			
Smoked	858 (50%)	Weekend light exercise	282 (16%)			
Never smoked	840 (49%)	Moderate activity 3x per week	400 (23%)			
Don't know, no response	14 (1%)	Heavy activity 3x per week	89 (5%)			
<b>Current or prior medical conditions reported at baseline</b>		Heavy activity 5x per week	78 (5%)			
<i>27 of 34 solicited medical conditions, listed by descending frequency</i>		<b>Medications, vitamins, supplements at baseline</b>				
High blood pressure	971 (57%)	Median (25 <sup>th</sup> , 75 <sup>th</sup> ) reported	7 (4, 11)			
High cholesterol	916 (54%)	10+ reported, n (%)	583 (34%)			
Rheumatoid arthritis	883 (52%)	<b>Top 5 reported medications</b>				
Obesity	605 (35%)	Lisinopril	342 (20%)			
Depression	592 (35%)	Hydrochlorothiazide	279 (16%)			
Osteoarthritis	530 (31%)	Omeprazole	277 (16%)			
Diabetes	421 (25%)	Levothyroxine	240 (14%)			
Asthma	299 (17%)	Metformin	237 (14%)			
Thyroid disease	269 (16%)	<b>Samples in inventory, collected at baseline</b>				
Skin cancer, not melanoma	207 (12%)	<b>Sample</b>	<b>Container, Size</b>	<b>Participants</b>	<b>Aliquots</b>	<b>Freezers</b>
Osteoporosis/Osteopenia	196 (11%)	Plasma	Cryovial, 0.5 mL	1,579	21,461	0.378
Heart attack or angina	171 (10%)		Cryovial, 4.0 mL	0	0	0
Other autoimmune disease	163 (10%)	Serum	Cryovial, 0.5 mL	1,590	14,270	0.251
Coronary artery disease	160 (9%)		Cryovial, 4.0 mL	0	0	0
Emphysema or "COPD"	157 (9%)		Cryovial, 5.0 mL	1,420	1,421	0.050
Gout	157 (9%)	Whole blood	PAXgene RNA	1,520	3,486	0.203
Atrial fibrillation	122 (7%)		Vacutainer, 2.0 mL	814	1,295	0.037
Other mental illness	113 (7%)		Vacutainer, 3.0 mL	0	0	0
Stroke	108 (6%)		Vacutainer, 4.0 mL	0	0	0
Congestive heart failure	76 (4%)	Buffy coat	Cryovial, 2.0 mL	1,161	1,162	0.020
Kidney disease	75 (4%)	Urine	Cryovial, 4.0 mL	0	0	0
Other type of cancer	75 (4%)		Cryovial, 0.5 mL	6	6	0.0001
Multiple sclerosis	67 (4%)		Cryovial, 10.0 mL	1,538	4,924	0.390
Melanoma	56 (3%)	Total				1.3291
Breast cancer	49 (3%)					
Liver disease	49 (3%)					
Crohn's disease/ulcerative colitis	48 (3%)					

## MURDOCK Study participants with Rheumatoid Arthritis, N=1,712

### Participant status and data from MURDOCK Study follow-up surveys and electronic health records

#### Participant vital status

Alive	1,409 (82%)
Deceased	303 (18%)

#### Current Age

Median (25 <sup>th</sup> , 75 <sup>th</sup> )	68 (59, 77)
Min, Max	25, 90+

#### Follow-up metrics, study participation

Median (25 <sup>th</sup> , 75 <sup>th</sup> ) months since enrollment	130 (109, 142)
Median (25 <sup>th</sup> , 75 <sup>th</sup> ) years since enrollment	11 (9, 12)
Median (25 <sup>th</sup> , 75 <sup>th</sup> ) yearly follow-ups complete	6 (3, 9)
Overall completeness of follow-up, n/N (%)	9,380 / 14,016 (67%)
At least one (1) follow-up survey complete, n (%)	1,565 (91%)
100% completion (n, %)	530 (31%)
Last completed follow-up ≤ 18 months	883 (52%)
Enrolled in one or more other studies	754 (44%)

#### Available EHR datasets by source (any ICD code)

Any source	834 (49%)
Novant Health	569 (33%)
Cabarrus Health Alliance	300 (18%)
Cabarrus Rowan Community Health Centers	123 (7%)
Bethesda Health Center	30 (2%)
Community Free Clinic	20 (1%)
Atrium (Carolinas Healthcare)	0

#### Available EHR data domains

Diagnoses	834 (49%)
Labs	652 (38%)
Vitals	516 (30%)
Medications	626 (37%)
Allergies	334 (20%)
Immunizations	260 (15%)
Problems	455 (27%)
Procedures	331 (19%)
Hospitalizations	282 (16%)

#### Insights from available EHR data

Date range: June 1993 (first encounter), Jan. 2021 (last encounter)	
Number of days between first and last encounter:	
Median (25 <sup>th</sup> , 75 <sup>th</sup> )	1,696 (192, 3,043)
Min, Max	0, 10,034

#### Select phecodes, mapped from diagnosis codes

Phecode	Description	Group	n, ppts
401.1	Essential hypertension	circulatory system	243
272.1	Hyperlipidemia	endocrine/metabolic	209
250.2	Type 2 diabetes	endocrine/metabolic	111
530.11	GERD	digestive	104
261.4	Vitamin D deficiency	endocrine/metabolic	85
278.1	Obesity	endocrine/metabolic	79

#### Select laboratory tests

Test	Labs	Participants
Comprehensive metabolic panel	2,421	366
CBC and differential	1,814	318
Basic metabolic panel	1,557	278
TSH	990	268
Hemoglobin A1C	1,402	262
Lipid panel	1,139	254
CBC	1,389	235

#### New medical condition diagnoses reported in follow-up

15 of 34 solicited medical conditions, listed by descending frequency

Rheumatoid arthritis	821 / 829 (99%)
Osteoarthritis	437 / 1,182 (37%)
High cholesterol	244 / 796 (31%)
Osteoporosis/Osteopenia	223 / 1,516 (15%)
High blood pressure	217 / 741 (29%)
Depression	195 / 1,120 (17%)
Obesity	189 / 1,107 (17%)
Other autoimmune disease	158 / 1,549 (10%)
Emphysema or "COPD"	155 / 1,555 (10%)
Diabetes	150 / 1,291 (12%)
Thyroid disease	138 / 1,443 (10%)
Skin cancer, not melanoma	135 / 1,505 (9%)
Asthma	131 / 1,413 (9%)
Gout	128 / 1,555 (8%)
Kidney disease	120 / 1,637 (7%)

#### Procedures reported in follow up

CT or MRI scan	1,197 (70%)
Joint x-ray	1,092 (64%)
Chest x-ray	1,052 (61%)
Heart/cardiac stress test	611 (36%)
Joint replacement	344 (20%)
Heart/cardiac catheterization	212 (12%)
Heart/cardiac angioplasty or stent	149 (9%)
Coronary artery bypass surgery	71 (4%)

#### Hospitalizations reported in follow up

Participants reporting 1 or more hospitalizations	863 (50%)	
Unique hospitalizations reported	1,445	
Median (25 <sup>th</sup> , 75 <sup>th</sup> ) hospitalizations reported	2 (1, 3)	
Coded reasons for self-reported hospitalization listed in descending frequency	Events	Participants
Uncoded	1,088	568
Surgery	220	160
Knee replacement	149	110
Pneumonia	78	59
Chest pain	66	55
Fracture	64	55

#### Body mass index (BMI) at most recent completed follow up

<18.5 (underweight)	26 (2%)
18.5 - 24.9 (normal weight)	334 (21%)
25 - 29.9 (overweight)	506 (32%)
30+	691 (44%)

#### Medications, vitamins, supplements at most recent follow up

Median (25 <sup>th</sup> , 75 <sup>th</sup> ) reported	7 (4, 12)
10+ reported, n (%)	503 (29%)

#### Top 5 reported medications

Omeprazole	275 (16%)
Atorvastatin	268 (16%)
Lisinopril	267 (16%)
Levothyroxine	230 (13%)
Amlodipine	229 (13%)